STAT 411: Advanced Applied S181istics
Spring 2016
R.Guerra

Overview

This course is topical from semester to semester. This semester we will cover study design, modeling and some computational methods for inference,

Study design addresses the most important part of a statistical study, namely, how the data are collected. Indeed, the maxims, "bad data in, bad data auf' and "garbage in, garbage out," make clear that if careful attention is not given to how one collects data, then there is little to be done to obtain valid statistical results (e.g., statistical summaries, p-values, confidence intervals, model fits, etc.). As part of the story we also distinguish between observational studies and experiments; the former are popular in many disciplines (especially sociology and epidemiology), but they can only point to association while experiments may point to causation. Survey principles and sampling will also be part of the study design material.

A prerequisite to STAT 411 is linear regression. We will extend the usual fixed effects regression model to include random effects. We will also cover examples of the generalized linear model (GLM), including binomial and Poisson regression to model counts. For example, we may want to develop predictive models for a certain disease based on a number of predictors such as age, sex, and BMI.

The last topic of the semester concerns some inferential methods that are made possible by modern computing power. The bootstrap is a methodology that is typically used to estimate standard errors or confidence intervals, especially in nonparametric settings. The Expectation-Maximization (EM) algorithm is a method for finding maximum likelihood estimates and their standard errors. Permutation methods are useful for generating reference distributions in hypothesis testing. When many hypothesis tests are evaluated, traditional error rates can be misleading and multiple testing corrections can be applied to obtain more valid decisions.

Prerequisites: STAT 310 *Probability and Statistics* and STAT 410 *Unear Regression.* Familiarity with matrix analysis or linear algebra is strongly recommended but not required. However, the course will make use of aspects of linear algebra that were covered in STAT 410. Students without 310 and 410 may not take the course.

Textbooks: Given the nature of the course, *there are no required textbooks.* Notes or references to course material will be given throughout the semester. For reference, the following are some standard textbooks that cover various parts of the course, but they are not required for the course.

Applied Statistics

Freedman, Pisani, and Purves (2007), *Statistics,* 41h ad., Norton. The best book Iknow of on the nature of statistics and that teaches you how to think like a statistician. Any edition will do.